Journal of Urban Economics 85 (2015) 16–33

Contents lists available at ScienceDirect

Journal of Urban Economics

www.elsevier.com/locate/jue

Identification in a model of sorting with social externalities and the causes of urban segregation

Maximilian Kasy *

Department of Economics, Harvard University, Littauer Center 200, 1805 Cambridge Street, Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Received 4 August 2013
Revised 9 October 2014
Available online 18 October 2014

JEL classification:
C31
D62
R21
R23

Keywords:
Identification
Sorting
Social externalities

ABSTRACT

This paper discusses nonparametric identification in a model of sorting in which location choices depend on the location choices of other agents as well as prices and exogenous location characteristics. In this model, demand slopes and hence preferences are not identifiable without further restrictions because of the absence of independent variation of endogenous composition and exogenous location characteristics. Several solutions of this problem are presented and applied to data on neighborhoods in US cities. These solutions use exclusion restrictions, based on either subgroup demand shifters, the spatial structure of externalities, or the dynamics of prices and composition in response to an amenity shock. The empirical results consistently suggest the presence of strong social externalities, that is a dependence of location choices on neighborhood composition.

1. Introduction

Urban areas in the United States and across the world show large degrees of social segregation across neighborhoods. A large and rising degree of segregation of immigrant groups across neighborhoods in US cities since 1920 has been documented, for instance, by Cutler et al. (2008). This is of concern if the social environment in neighborhoods is an important determinant of life outcomes. There are two polar explanations of segregation. Households might sort across locations because of different willingness to pay for exogenous location characteristics, which may be due to differences in income or differences in preferences. This is the explanation emphasized by accounts of sorting such as the classic Tiebout (1956) and Rosen (1974) models. Alternatively, households might care about who their neighbors are, and hence choose their neighborhood based on demographic composition. This possibility was discussed by Schelling (1971) and Becker and Murphy (2000).

The present paper discusses identification problems arising in models which allow for both possibilities. In the setup considered in this paper, households have to choose whether or not to locate in a given neighborhood based on exogenous neighborhood characteristics, and based on the endogenous composition of the residents of a neighborhood. We say social externalities are present if demand depends on endogenous composition. The local housing market is in equilibrium if the composition of households that want to locate in a neighborhood equals the composition of those that are in the neighborhood, and if total housing demand equals housing supply. This setup builds on several important recent contributions to the urban economics literature, in particular Bayer et al. (2007) and Caetano (2009). These authors estimate discrete choice models of sorting that recognize the possibility of a preference for neighborhood composition.

The central contribution of the present paper is to provide a discussion of nonparametric identification in this context. The main goal is to empirically distinguish between the two explanations of segregation, and in particular to test whether social externalities are empirically relevant. It is shown that without further restrictions the presence and degree of social externalities are not identified. This is because, in equilibrium, both composition and rental prices are functions of the exogenous neighborhood characteristics. This prevents the separate identification of the effect of either on demand. Identification requires exclusion restrictions that generate independent variation of composition and the exogenous arguments of demand. Several types of such exclusion restrictions are discussed here. In the setup analyzed, no restrictions on

* Address: Department of Economics, Harvard University, Littauer Center 200, 1805 Cambridge Street, Cambridge, MA 02138, United States.
E-mail address: maximiliankasy@fas.harvard.edu

http://dx.doi.org/10.1016/j.jue.2014.10.003
0094-1190/

functional forms or the nature of heterogeneity of households or neighborhoods are imposed. Discussions of nonparametric identification have been fruitful in the development of many applied fields in recent years, see for instance Manski (2003) or the review in Matzkin (2008).

The presence of social externalities in sorting is of relevance for several reasons. First, it poses a methodological problem in the estimation of willingness-to-pay parameters, which in turn are often used for cost-benefit analyses of policies. Second, externalities matter for understanding the causes of social segregation across locations and can amplify the effects of policies on segregation. Third, if externalities are strong, multiple equilibria in population composition at a given location arise. Multiple equilibria in turn can imply discontinuous and large effects of demand shifting policies, as emphasized by Schelling (1971) and Card et al. (2008). Finally, it is interesting to contrast the importance households attach to neighborhood composition in their location choice with the available evidence on the effect of neighborhood environment on observable outcomes. Evidence on the latter is mixed, see for example Katz et al. (2007). The present paper, on the other hand, finds strong effects of composition on location choice. These results are of course consistent with each other, but suggest that households care about neighborhood composition for other reasons than the causal impact of neighborhood composition on observable outcomes.

Three possible solutions to the identification problem are discussed in this paper. The first approach uses exogenous shifters of demand of certain subgroups that are excluded from the demand of other subgroups. Such shifters allow one to construct instruments that affect neighborhood composition, without directly affecting the demand of some subgroups. Using such instruments we can estimate the causal impact of composition on demand of these subgroups. This builds on the idea of randomized subgroup treatment used in the identification of peer effects, as recommended in Moffitt (2004) and applied for instance by Duflo and Saez (2003).

The second approach exploits the spatial structure of cities in an extension of the baseline model, allowing for interactions across adjacent neighborhoods. Identification comes from the assumption that exogenous demand shifters for neighborhoods beyond a certain distance are excluded from local demand. This allows one to use demand shifters for neighborhoods at a certain distance as instruments which affect local composition through their impact on the composition of intermediate neighborhoods, without directly affecting local demand. Using such instruments we can estimate the causal impact of composition on demand of all subgroups, as well as the causal impact on housing prices. The latter measure the marginal willingness to pay for housing in the neighborhood. This idea is analogous to the use of social network structures to identify endogenous versus exogenous peer effects, as in Bramoullé et al. (2009) and De Giorgi et al. (2010).

The third approach is based on a dynamic extension of the baseline model which is discussed in more detail in the supplementary appendix. This dynamic extension assumes search frictions in moving from one neighborhood to another. The third approach uses the finding that, under certain conditions, past amenity shocks are excluded from future price changes because the value of amenities is immediately reflected in rental prices. Composition, however, does adjust with delay due to search frictions, and hence prices adjust to this composition change with the same delay. Past amenity shocks can therefore be used as instruments which affect future composition changes without directly affecting future changes in prices. Using such instruments we can estimate the causal impact of neighborhood composition on housing prices (marginal willingness to pay). The dynamic model considered is similar to search models of the labor market as surveyed in Pissarides (2000). It builds upon search models of the housing market such as Wheaton (1990).

These approaches are applied to data from the Neighborhood Change Database (NCDB), which aggregates US Census data to the level of census tracts. The composition variable considered is Hispanic share. Various instruments for neighborhood composition are constructed that build on the three approaches to identification just discussed. All instruments yield surprisingly consistent estimates. They suggest that a 1% increase in the Hispanic share of neighborhood population results in a 6% to 10% decline in non-Hispanics' demand, and a 3% to 4% rise in Hispanics' demand. Housing prices appear to decline by around 0.5% to 1% for a 1% increase in Hispanic share. These results are also consistent with the conclusions of Cutler et al. (2008), who use variation in segregation across time, city, and immigrant groups in trying to disentangle the causes of segregation. One might wonder why we are focusing our main empirical analysis on Hispanic share, rather than on other dimensions of urban segregation. The main reason is that immigration created a lot of arguably exogenous variation in composition, which we exploit.

The model in this paper is described in terms of households choosing a neighborhood and paying rents. However, most of the insights should apply to other contexts of sorting. Examples include sorting of workers across firms, students across schools, customers across mobile-phone network providers, faculty across universities, or the spatial agglomeration and dispersion of firms. In each of these settings agents might have a (reduced form) preference for peers, which is empirically hard to separate from location heterogeneity, but which has implications for interesting counterfactuals.

Some further relevant contributions in the recent literature have to be mentioned before proceeding. Solutions to the omitted variable problem in hedonic or choice regressions have been proposed by Black (1999), who controls for border fixed effects, and by Chay and Greenstone (2005), who use exogenous variation in amenities. Nesheim (2001) and Graham (2008b) discuss identification issues in specific models of sorting where peer composition enters an educational production function. Heckman et al. (2002) and Ekeland et al. (2004) derive identification of preferences from cross-sectional price data based on functional form restrictions (separability). Chiappori et al. (2009) show the equivalence of hedonic sorting, matching and optimal transport problems and derive existence results for equilibria in these models.

The rest of the paper is structured as follows: Section 2 introduces a model of locational sorting and discusses its assumptions and the fundamental identification problem in this model. Section 3 proposes three solutions to this problem, based on subgroup shifters, the spatial structure of cities, and the dynamic structure of neighborhood composition and prices. Section 4 applies these three solutions to the NCDB data. Section 5 concludes. All proofs are relegated to Appendix A. Additional discussions can be found in a supplementary appendix.

2. Model and identification problem

This section will first state the model assumptions and the basic non-identification result which motivates the present paper. Then the model assumptions will be discussed. A special case of the model will be used to provide some graphic intuition for the comparative statics of the model and for the source of the identification problem.

We will consider the following model of the local housing market in a given neighborhood. There are $\mathscr{C}$ types of households, $c = 1, \ldots, \mathscr{C}$. A neighborhood is characterized by (i) the mass (number) of households of each type, $M = (M^1, \ldots, M^{\mathscr{C}}) \in \mathbb{R}^{\mathscr{C}}$, (ii) a (rental) price P, and (iii) an exogenous vector $X \in \mathbb{R}^{k_X}$ of *all* other location characteristics and factors influencing demand or supply. An example component of the neighborhood characteristics vector